WRF Performance/Scaling

Dr. Dimitris Dellis

GRNET

Online 28 Apr. 2025

WRF Performance/Scaling

- ► ALL CPU codes speed depend on CPU speed.
- ► CPU speed depends on vectorization features.
- ► One need to apply specific compilers cpu optimization flags to activate if possible the use of these features.
- ► WRF default values are : -O2 with GNU compilers, -O3 with Intel.
- ► Makes this difference ?

WRF Performance/Scaling

► 6 Hours forecast over Greece, same period, 1 node, 20 cores

GNU	-O2	1284
GNU	-O3 -mavx2 -mfma	1189
Intel	-O3 -xCORE-AVX2	831

WRF Performance/Scaling

- ► WRF is memory bound $\implies$ needs high memory bandwidth.
- ► WRF use 2D grid $\implies$ need fast (low latency) communication between procs. Intranode MPI communication use shared memory.
- ► Example : 3 Identical machines (vendor, CPU, cores, amount of memory) had dramatically different performance, Ratio of speed 1:0.55:0.3
- ► After testing, 1 had 8x16 GB DIMMS, 1 had 4x32 GB DIMMS, 1 had 2x64GB DIMMS. Useful when you buy hardware for use with memory bound applications.
- ► Automatic grid decomposition is based on the number of available/requested MPI tasks. Number of tasks must be appropriate to obtain good 2D grid decomposition.

WRF Performance/Scaling

► Scaling : Execution Time vs number of Nodes/cores

| Nodes	Cores	Grid	Time [s]	Efficiency[%]	coreHours
1	20	4 x 5	1408	100	7.82
2	40	5 x 8	752	93.6	8.35
3	60	6 x 10	490	95.7	8.16
4	80	8 x 10	406	86.7	9.02
6	120	10 x 12	302	77.7	10.06
8	160	10 x 16	259	67.9	11.51
10	200	10 x 20	212	66.4	11.77
12	240	15 x 16	213	55.0	14.20
14	280	14 x 20	185	54.3	14.39
16	320	16 x 20	178	49.4	15.82

WRF Performance/Scaling

► Are you ready to pay 16x cost to get 7.9x speed up ?
  ► Makes sense for operational runs - you need the forecast 1-2 times per day at certain time.
  ► For modelling where you need to complete a large period, it is better to run with lower number of cores.
  ► With these numbers a choice of up to 120 cores is reasonable.
  ► Remember, cH may mean Euros.

WRF Performance/Scaling

▶ Computational Inhomogeneity.
▶ Profile. Measure the percentage of time each MPI task just waits for data from its neighbors. Color indicates the percentage in % of time. Range is 30 - 90 %.
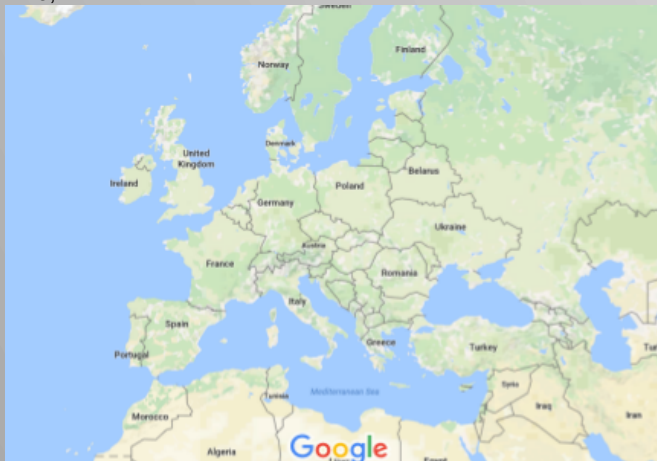▶ WRF run with 900 cores (grid 30x30). Axis labels indicate X and Y index in 2D grid.

WRF Performance/Scaling

▶
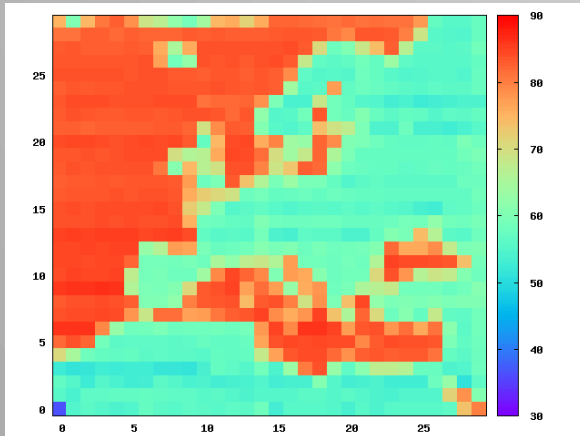
▶ Says something ?

WRF Performance/Scaling

► if no,

Thanks!

EuroHPC

EuroCC2 , Online, GRNET, 28 Apr. 2025

eurocc-greece.gr